

教育評価の理論と実践 —真正の評価をめざして—

山口 陽 弘・石 川 克 博

群馬大学大学院教育学研究科教職リーダー講座

Theories and Practices of Instructions for Educational Evaluation : Aiming for Authentic Assessment.

Akihiro YAMAGUCHI, Katsuhiko ISHIKAWA

Graduate School of Education, Program for Leadership in Education

キーワード：真正の評価、教育評価、理論、実践

Key words : authentic assessment, educational evaluation, theory, practice

(2011年10月31日受理)

群馬大学教育学研究科専門職学位課程（以下、「教職大学院」）では平成20年度の発足以来、ほぼ全ての授業を研究者教員と実務家教員のチーム・ティーチング（以下TT）で実施している（佐藤他，2011）。本稿の筆者2名は「教育評価の課題と実践Ⅰ」「同Ⅱ」を、平成20～23年度の4年間にわたってTTで担当してきた。

この4年間で行ってきたことを集約すると、山口が科学的な心理学の視点や統計学的な発想に基づいて、「理論」的な教育現場へのヒントを提示してきた。それに対して、石川が学校現場では、実際に求められている教育評価の「実践」的な具体例を紹介し、やはり直接的に教育現場へのヒントとなることを提示してきた。それらのヒントを受講生に投げかけることで、受講生自身がこれまで実践したり考えてきた教育評価を振り返り、改善してもらおうというリフレクションを促す授業であり、それは現職教員及び教員志望の院生にとって有効なものであったと考えている。

こうした教育評価についての授業実践、すなわち研究者教員と実務家教員とのTTを通じて分かった事がある。「理論」と「実践」とは一見かけ離れているように思われがちであるが、その目指す方向や考え方に関

しては驚く程一致するという点である。その一致点とは、研究者も実務家も、最終的に教育評価に関しては「真正の評価（authentic assessment）」を目指すという点である。

この真正の評価とは、1980年代後半からアメリカの教育評価の歴史において出現した概念であり。近年日本でも教育評価にその考え方が導入されつつあるものである（e.g.田中，2005）。本稿では、この真正の評価とは何か、また真正の評価を目指すには具体的にどうすればよいかという点を中心に論じていく。

1. 真正の評価とは何か

1. エバリュエーションとアセスメント

教育評価については、「エバリュエーション（evaluation）」と「アセスメント（assessment）」という二つの観点で整理して論じられる事が多い（e.g.ドラム他，2007）。

もちろん、両者を厳密に区別せずにほぼ同義とする立場もあるが、通常、両者は異なった意味で使われる。その異なった意味を前提として、両者の長所を併せた「結び」の部分で「真正の評価」と定義し、両者の長

所を活かしていこうとする試みが、近年の日本の教育界でも受け入れられており、本稿でもその定義を採用することにする。

まず、両者のニュアンスを対比的にまとめておく。「エバリュエーション」というのは、どちらかといえば児童生徒を教師が「上からの視点で」評価し、価値判断を「下す」というニュアンスがある。それに対して、「アセスメント」は多角的な視点で教師が情報収集して、児童を診断するものであり、次の教育活動に向けて改善する方策を打ち出すための行為であるというニュアンスがある。

今述べた「エバリュエーション」のニュアンスでは、「エバリュエーション」にマイナスの意味しかないようであるが、そうとも言えない部分がある。価値判断は教育活動の中では避けられない重要な側面である。また、その価値づけがあるゆえに児童生徒は勉強しようとする動機づけられるプラスの側面もある。そもそも「エバリュエーション」という言葉を最初に用いたタイラー、R. W. (1978) は、テストのために出題される内容を学生たちが勉強しようとする、その強力な影響力を、教育実践の改善に役立てないかと考えていたのである。

そこに「アセスメント」としての診断的な部分や、さらなる教育活動に向けての改善の面がさらに導入する発想で生まれたのが「真正の評価」という概念なのである。

これはブルーム他 (1973) の「形成的評価」や、「指導と評価の一体化」ということとほぼ一致する概念であり、古くて新しい評価の問題に解決を与えるのが「真正の評価」であるように思われる。

2. 真正の評価の問題点

複数の論者の真正の評価の定義を借用して、さらに議論を進めよう。真正の評価とは、「大人が仕事場や市民生活、個人的な生活の場で試されている、その文脈を模写する事」と定義される (Wiggins, 1998)。また、シャクリーほか (2001) は、真正の評価とは、「リアルな課題」に取り組ませるプロセスの中で子どもたちを評価することであると定義している。

この二種類の定義は、ほぼ同じことを述べていると思われる。また、これは心理学でも認知革命以降で重要視された生態学的妥当性 (ecological validity) (ナイ

サー, 1989) という概念とも合致するであろう。すなわち、実験室的条件で得られる心理学的法則は、非常に特殊な状況下での法則であり、現実にはありえない。したがって、できるだけ日常文脈に即した、生態学的に妥当な状況で心理学実験は行われ、そこでの法則こそが意味あるものであるという主張である。以上をまとめれば、教育評価を「真正な」ものにするためには、より日常文脈に即した方法でリアルな課題に取り組みせる必要があるということである。

このように評価を再定義し、真正の評価へと高めていくことが、教育評価についての最大の目的であり、課題であることに異論がある論者はいないであろう。

しかし、その真正の評価の中身とは一体何なのだろうか。あるいは具体的に真正の評価をするためにはどうすればよいのか。この言葉に内実を与える事は、そう容易なことではない。というのは、真正の評価とは要するに「よい評価」を目指すことを目新しい言葉で言い換えているに過ぎず、そこに実体を与える方策についてはまだまだ未検討、不十分であるからである。

そこで山口は以下、IIからIIIにおいて、科学的、統計学的、心理学的な視点で、そのためのヒントとなりそうなことを列挙して、真正の評価に近づくための方策を論じている。

一方石川は、実践的な視点 (特に実際に教える教材やPISA調査などの具体的な設問) からやはりヒントとなりそうなことを論じる事で真正の評価に近づくための方策をIVからVにかけて論じている。

この両者の検討は、理論と実践という異なった観点から、真正の評価という概念に実体を与える試みである。ここで注意を喚起したいのは、真正の評価という概念に内実を与えるためには、理論だけではもちろん不十分だが、同様に実践のみでも不十分であるという点である。理論と実践の往還こそが、「真正の評価」を豊かにさせ、教育活動を具体化することであると考えられるものである。

II. 教育評価に及ぼす観察者、実験者効果

1. ホーソン効果

社会心理学、産業心理学の古典として、ホーソン研究と言われる有名な一連の実験がある。これは1924～

1932にかけて、シカゴのウエスタン・エレクトリック会社ホーソン工場で、二十代前半の女子工員数名を対象に行われた実験である (c.f. 吉田, 1969)。研究主催はMayo, E.であり、いかなる作業環境が生産性を向上させるかを実験的に検証しようとしたものである。その結果、さまざまな作業環境を改善するよりも、自分たちが「選ばれた者である」という意識の方が「勤労意欲」を著しく高めるということが発見された。

本来ならば比較対照群として、実験群よりも作業環境が劣ると予想された群ですら、作業成績は高まっていったのである。つまり、人間は一定の条件では一定の働きしかできないという仮説が崩され、また公式的な組織・人間関係よりも、非公式的な組織・人間関係の方がはるかに人間に影響を与えることがわかった点でも、画期的な研究であった。ここで見いだされた効果のことをホーソン効果 (Hawthorne effect) と呼び、観察者が実験に与える効果を示す一例とされている。

これは、特に教育場面で生じやすいことであり、教師が教室場面に置き換えて考えねばならぬ重要な問題であろう。たとえば、研究授業で、担当教諭以外に、校長、教頭、同僚、場合によっては保護者などが授業参観に来る場合がそれに該当する。これはまさに、典型的なホーソン効果が生まれる場面ではあろう。多くの見知らぬ人から観察される事で児童生徒は「選ばれた者 (クラス) である」という感覚を持つだろう。そこまではいかなくとも、ふだんよりは一般に学習意欲は高まるであろうし、少なくとも努力しようという気持ちにはなるであろう。これらの観察者による諸々の効果が、実際のその研究授業の実践手法よりも、はるかに強力である場合が多分にあるように思われる。すなわち、当該の教育技法が直接の要因になったのではなく、観察者効果が児童生徒の「学習意欲」を高めた結果、教育効果が生まれるということである。

もちろん、児童生徒が学習意欲を高めること自体は望ましいことである。しかし、研究授業での教育実践の技法を、自らの教育現場に持ち込む際には、ホーソン効果があることを想定し、その効果のある程度割り引いて、特定の教育技法を神格化しないように意識する必要があることを注意したいのである。

2. クレバーハンスの事例

同じく、心理学上の古典として、クレバーハンスの

話も示唆に富む。これは、二十世紀の初頭、計算ができるように調教されたハンスという馬の話である (c.f. 鈴木, 2008)。

実際にハンスは計算能力があったわけではなく、問題の出題者の反応をみて、期待されるように振る舞っていただけであったというものである。つまり、「実験する側の人間が知らないうちに被験者に答えや反応の手がかりを与えてしまう場合があるということ」(鈴木, 2008) あり、こうした実験者の影響を「実験者効果」と呼び、特に動物実験の際には、「クレバーハンス効果」とも呼ばれ、心理学実験の際に、こうした効果が入り込まぬようにする事が、鉄則とされている。そのために、二重盲検法 (double-blind method) という手法が厳密には求められる事になる。

この二重盲検法とは、被験者のみならず、実験者もその問題や解答が分からないようにしておいて実験を実施するという手法である。先のクレバーハンスの例で言えば、算数の問題を出題した人が、問題自体を知らず、したがって、解答がわからぬ状態にすることである。このとき、ハンスはどう答えていいか分からず、計算ができなかったのである。これは新薬の効果を検証する際に、偽薬効果を消すために求められる検証法でもある。

しかし、教育場面において、教育者が問題の解答を知らず、さらに問題自体も知らずに出題するという事態があるだろうか。もし、あるとすればそれは、せいぜいセンター試験の試験監督を大学教員がするときぐらいの状況であろう。まさに、「非教育的」で、評価を「下す」場面で、試験監督以外の行為を教員がしなくてよい場面である。

実際には、教師は児童生徒にある行為を強く期待して教育活動を行っている。少しでも思うような行為が出れば、言語的・非言語的、意識的・無意識的な正のフィードバックを返している。この手がかりをもとに児童生徒はしばしば回答を決定する。このような場面で教育評価がなされる事が、評価を難しくさせているのだと思われる。つまり、本当に問題を理解した上での解答なのか、教師の反応をみてそれに反応してたまたま正解しただけなのかを、教師は自身の期待を脇に置いた上で峻別する必要があるのである。

優れた教師は、時にあえて児童生徒に疑問を投げかけるような、またそれまでの文脈とは逆の問いかけを

して、より深い理解を促す事がある。誘導尋問的に望むような答を出させて、浅い理解で授業を済ませるのではなく、深い理解へと導くために、一見授業の流れからすると、本来の回答とは逆の回答を喚起する発問系列も、時には必要であろう。つまり、教師の期待に沿った回答を求める発問系列のみでは、教育とはならない。児童生徒をクレバーハンスにすることが、教育の本来の目的ではないのである。

3. ピグマリオン効果、ハロー効果

2で述べたクレバーハンス効果と類似した効果として、ピグマリオン効果 (pygmalion effect) (予言の自己成就効果、教師期待効果、提唱者の名前を取ってローゼンタール効果とも) と言われるものがある。これは Rosenthalら (1979) が1964年に見いだしたもので、小学校1～5年生全員と入学前の幼稚園児に、近い将来の知的能力を予測できる (と教師には説明するがこれはウソ) 知能検査を実施し、その結果から、将来の伸び (これは全く無作為に指摘する) を教師に告げたところ、新学期開始後8ヶ月後に伸びると言われた児童が顕著な伸びを示したという実験である。

この実験は原典を読むと、実験手続きに不備もあり、また倫理的問題もあり、現在となってはあまり信憑性は置けないものとされている。しかしやはりクレバーハンスの事例のように示唆に富むものであり、特に教育的な場面では、教師が考えねばならぬ深い問題を抱えている。つまり、評価以前に、よい期待をするという間接的な効果が、児童生徒をよい方向に導く事もあるということである。いわんや教育評価そのものの直接的効果が、教師にも児童生徒にも与える影響は、はるかに強力なものであろう。先にも述べたタイラーの主張した、エバリュエーションの持つ強力な影響力の一つである。これを教師は忘れてはならないだろう。

同じく、ハロー効果 (halo effect) という問題も、教育評価の際には、注意すべきだとしばしば指摘される。光背効果、後光効果とも言われるもので、人物評価の際に、特定の側面で望ましい (望ましくない) 特性があったとき、それに引きずられて他の側面も同様に望ましい (望ましくない) と見なしてしまうことを指す。

これなども、教育場面ではよくあることであり、特に小学生相手の場合、基本的には全科目を教師は担当するため、たとえば算数ができる児童を、理科もでき

ると即断したりすることは注意せねばならないだろう。他にも、ある時点での小テストの成績が良かったり (悪かったり) したとき、次の小テストの判断の際に同じような見方をして採点、評価することはないだろうか。これも一種のハロー効果と言える。

児童の見取りをするときに過去に関する情報を使うことは悪いことではなく、重要なことである。しかし、教育評価を歪ませないためには、過去の情報からいったん距離を置いてその時点での評価をすべき時があるだろう。

要するに人間は怠惰であり、認知的に経済的な見方をしがちであり、よほど注意深くないとステレオタイプ的な見方をしてしまうことが、数々の認知心理学や進化心理学の知見から得られている (e.g. ギロビッチ, 1993)。それらの錯誤や誤謬を少しでも乗り越える事で、正確な認識に達するということである。ここまで述べてきたような、人間が起こしがちな錯誤に関する□□効果と言われる様々な心理学的知見は、「真正の評価」を目指すヒントに十分なりうるであろう。

4. 教育臨床的な観点からの提案

上述の1～3までの各種の反応の歪曲や教師の判断のミス、いかにすれば防ぐ事が可能であろうか。結論を先に述べれば、魔法のように簡便な方法はないであろう。逆に言えば、各種の心理学的知見によれば、人間は上記のように間違えやすいことは実に頑健な現象なのであると、証明されているのである。また、歪みのない完全な客観性を教育評価に与えることは、カントを持ち出すまでもなく、最初から不可能とも言えることである。これは心理学というものが、そして教育学というものが人間を相手にした学問であるゆえ、やむをえないことである。

しかし、山口は、次の三つの枠組みが、教育評価を可能な限り補正し、歪ませないために参考になると考えている。

第一がインフォームド・コンセント (同意) である。つまり、教育評価を行う際に、評価される事自体を、児童生徒が拒否、あるいは同意していない状況では、そもそも評価は不可能である。テストの実施を、児童生徒が納得することが教育評価の必要条件である。もちろん、小学生にテストの必要性を認識させることは容易ではない。実際には強制的にテストが実施される

事はあるだろう。しかし、そこでテストはなぜなされているのかということについて、各発達段階に応じて、児童生徒が納得していなければ、そこで教育評価は、歪曲されてしまうであろう。

第二がフレーミング（枠組み作り）である。教育評価のために様々な質問やテストを行ったりすることがある。このときに、質問者の意図を汲まずに、あるいは意図を越えて回答した場合、その回答は歪んだものとなる。たとえばキャランドラが作った、有名なたとえ話がある。これは、気圧計を使って、ビルの高さを測るように求められた学生が、本来の気圧計の使い方以外の用法（たとえば気圧計をビルの管理人に持って行って、その代償にビルの高さを教えてもらうとか）を使って回答したために、物理学の試験に落ちてしまうという創作されたたとえ話である（池田，1992）。

つまり、今質問されていることが、どういう枠組みで聞かれているか、また、それにどういう枠組みで答えればよいかを児童生徒が把握していることが、教育評価の際の必要条件なのである。人間の知というものが機械の知（人工知能）と最も異なる点は、こうしたフレーミングを、状況や場面に応じて自在に行うことであるとされる（e.g.辻井・安西，1988）。また、人工知能における最大の問題がこの「フレーム問題」であり、フレーミングが、自在にできるようにすることこそが、「人間の」教育であるとも言える。

他の場面でもたとえば、外国籍児童がテストへの回答ができないときには、日本特有の問題形式、つまり問題文のもつフレーム（枠組み）自体がよく理解できないから回答できない場合が往々にしてある。彼らに対してはフレーム部分の補償的な教育指導が必要である。逆に言えば、教育の中でなされるべき真の内容とは、フレーミング能力であり、それを評価すべきであるとすら言える場合もあるだろう。これは、個別の知識獲得を問うのではなく、知識獲得の枠組み自体を問うという、PISA型の学力を求めることにも繋がる問題でもあろう。

第三がラポール（信頼関係）である。これは教師と児童生徒との間の信頼関係である。あまりに当然のことなのでここで改めて書くまでもないことだが、教師への不信感が児童生徒にあったり、逆に教師が児童生徒に不信感を持っているときに、教育評価は真正なものから大きく歪むであろう。

以上の三点は、然るべき学級経営の上に立たねば成立しない事である。教育評価は教科教育が成立していなければ成立しない。しかし同様に、教科教育が成立するためには然るべき学級経営が必要である。学級経営の際に求められる必要条件が、この三点ではなかるうか。

III. 教育評価における統計学知識の必要性

ここまでは、特に心理学的・科学的な見地から述べてきたが、特に統計学的な視点からも重要なことは数々ある。その中で教育評価の中で特に関連しそうなトピックに絞って論じてみることにする。

1. 回帰効果

教育評価において、ある時点での評価としばらくしてからの評価との間で相関を考えたときに、相関係数が1（あるいは-1）になることはまずありえない。特に非常に高い評価を受けた者に焦点を当てて考えると、相関係数が1でもない限りは、必ず次の評価は低くなるということになる。

しかし、これを教師の側からすると、せっかく高いレベルに達したのに、次の成績が落ちてきたということに対して、（場合によっては）誤った理由を与えて（帰属させると心理学では言う）、児童生徒が怠けてしまったのではないかと推論したりする事がある。あるいは逆に、非常に低い評価を受けた者は、次回も低いままであることは珍しく、相関係数が1でもない限り、必ず評価は高くなる。これも低い評価を受けたので頑張ったのではないかという推論をしまいがちである。

こうした誤った推論を「回帰の誤謬」と呼び、我々人間が推論する際に、注意せねばならない誤りであると言われる。教育評価を歪ませないために重要なヒントとなるものである。

2. 教育目標設定時の重回帰式についての注意

小学校高学年において、走り高跳びを評価する際に、実際の教育現場で次のような数式を用いて目標設定し、その到達率などを観察時の評価基準としているという（池田・蒲地，1987）。

「目標記録(cm) = 身長(cm) × 0.5 - 50m走タイム(sec) × 10 + 120 (cm)」

この数式は、そもそも重回帰分析と言われる手法で、数百人でのデータをもとにして算出している。

この数式を元にする、身体の大きさや、50m走のタイムといった個人差要因を考慮した上で、個人内での目標設定が可能となる。その結果、単なる印象評定や絶対評価を越えて、個人内でどの程度の伸びがみられるかを判断することができる。確かに従来よりも工夫された、有効な目安となるべき指標であると思われる。

しかし、注意すべき点は、1で述べた回帰効果の問題である。たとえば運動能力がきわめて高いレベルの児童はそれ以上の伸びが難しい。逆にきわめて低いレベルの児童は伸びは簡単に得られる。場合によっては、上記の重回帰式よりも落ちたり（伸びたり）することが、別に不思議ではない。まさに回帰式によって導かれたものであるために、回帰効果が生じるのである。

また、重回帰に関しては予測率を明記しておかないと、現場の教師が上記の数式を絶対視し、そこから外れた児童に対して、誤った評価をする可能性が生まれる。この元となるデータも、今から二十年以上前の児童のものを元にして作成されている。2011年現在の児童を対象としたときに、このままの重回帰式が適切かどうかは、再検討の必要があるだろう。しかし、ごく最近でも、この回帰式をもとに評価基準が設定されているようである（藤田 他, 2010）。実際に2011年の山口・石川の授業である「教育評価の課題と実践Ⅰ」で、小学校体育を教えている現職教員から、この数式をそのまま利用して、各種の教育評価活動を行っているという報告を受けた。

もちろん、ほかに情報がないとき、また中程度の運動能力を有する児童にとっては、とりあえずの教育目標の設定としては一つの目安にはなるであろう。絶対評価から一歩進んだ評価や指導をしていくために、個人内での目標を設定していくことは有意義であり、そのために重回帰式を利用することが有効な一つの方法であることは確かである。要するに数式の絶対視を避けるためにも、重回帰に関する最低限度の統計的知識が必要であるということをも主張したいのである。

3. 信頼性と妥当性

山口が教育評価に関する文献を検討していて気づいたことがある。研究者に限定しても、教育評価に関して、おおむね二種類の背景が異なる論者がいる。それは教育学者と心理学者であるのだが、この両者は、教育評価を論じる際に、重視する点が大きく異なるのである。

心理学者はその多くが、統計的知識の必要性を説き、またその技術開発に余念がない。具体的にはテスト理論を軸に論を展開することを好む（e.g.池田, 1992）。教育学者はどちらかと言えばその問題の歴史や哲学を中心に論じることを好み、テストを作成するための技法（e.g.パフォーマンス課題）やカリキュラム・実際の教育場面との関係性を軸にして論じる傾向にあるようである（e.g.ギブス, 2001）。

山口自身は心理学を背景として教育評価を論じるため、本論文でも統計学的知識の必要性を説いてきたし、この重要性に関する認識に揺らぎはない。その統計学的知識のもっとも重要な核が、信頼性と妥当性であると考えている（信頼性と妥当性に関する詳細は、山口（2002）を参照）。

信頼性とは、教育評価に即して言えば、評価結果が安定している度合いである。誤差が評価に入ってくる割合が少ないほど信頼性が高いと言える。この信頼性に関しては、特に統計的見解からさまざまな有益な知見が得られている。

よく用いられるものが再検査信頼性やクロンバックの α 係数である。前者は、テスト（評価）を同一人に二回繰り返しても同じ値を示すとき、そのテストは再検査信頼性が高いと言われる。後者は、内的整合性、一貫性とも言われるもので、テスト項目間で整合性がとれていて、一貫性がある場合に、 α 係数は高くなり、優れた項目であると言われる。信頼性係数として、何をどのような目的で使うのかによるが、目安として最低でも0.7以上、できれば0.8以上の数値を得る事が望ましいと言われている。

一方、妥当性とは、真実性とも言われるもので、評価したい対象を、どれほど本当に評価できているかということを示す概念である。これは数値では示しにくく、テスト理論としてもその妥当性を確保するためのさまざまな技法は提案されてはいるのだが、決定的な方法はないと言える（これらのテスト理論の詳細は池

田 (1994) を参照)。

したがって、心理学者もどちらかといえばいかにして信頼性を上昇させるかという議論をしがちであるようである。しかし、心理学者も、もちろん信頼性よりも妥当性が重要であるという認識はしている。

たとえば心理学者は一般に血液型による性格判断を否定しており (e.g.大村, 1998)、山口自身も否定しているが、血液型という指標の「信頼性」に関してはきわめて高いことは認めねばならない。各種の心理検査の不安定性、信頼性の低さとは比べものにならない。A型の人が生涯A型であり続けるということは、測定論的に言えば実にありがたいことなのである。

しかし、血液型性格診断には、「妥当性」がない (少なくとも現時点ではないように思われる)。したがって、どれほど信頼性が高くとも妥当性がないため、テストとしては望ましくないのである。

では、教育評価という枠組みの中で、妥当性を上昇させるためにどうすればよいのだろうか。ここで再び「真正の評価」という概念が有効になってくる。信頼性を多少犠牲にしても、より日常的なリアルな文脈で、リアルな課題を解いてもらうことで妥当性を確保すること、その上で評価するというのが真正の評価の意図であった。これは妥当性を向上させる際にも、そのまま使える考え方であろう。教育評価をよりリアルであるような文脈や課題にすることが、妥当性向上のためにもっとも直接的な方法である。

しかし、「リアル」「リアリティ」というのは、わかりやすそうで、実はわかりにくい概念である。たとえばリアリティがあるというときは、それは本物でない対象に対してのみ言われる言葉である。映画や芝居や小説のようなリアルでないものに対して、リアリティがあると評価するのであって、事実に対してリアリティがあるということはあまりないであろう。むしろ事実は小説よりも奇であって、事実の方がリアリティがないようなことが往々にしてある。

したがって、ある教育評価に関して、それにリアリティがあるというのは、そもそも教育評価が実際の文脈からかけ離れたものであるということを前提としての批評なのである。

そのリアリティとはいったいどう考えればよいのだろうか。養老は、リアリティを訳すときに「現実性」と訳すのは誤りで、「真善美」と訳すべきだと述べてい

る (養老・玄侑, 2007)。すなわち、より真に、より善に、より美に近づくようになったときにはじめてリアリティが生まれるとしている。これは教育評価にリアリティを与えるために有効な指針を与えるものと考え

る。すなわち、教育評価に妥当性を与え、真正の評価にしていくためには、評価がなされる文脈が、本来の教育活動よりもさらに「真善美」に近づくような文脈と課題になることが必要なのである。それを全うすることは困難なことかもしれないが、たとえばPISAなどで使用されている、「良問」は、従来の特定の知識のみで解答できる問題よりは、あくまで相対的にはあるが、確かにリアリティがあるように思われる。評価に関するドラスティックな解決は難しいかもしれないが、教育評価も少しずつは改善されている。PISA型学力で出されている問題を吟味することから、各教員の教育評価活動にリアリティを与えて、真正の評価にすることが今、求められているのではないだろうか。

引用文献

- B. S. ブルーム, J. T. ヘスティングス, G. F. マデュス (著) 梶田叡一・渋谷憲一・藤田恵壘 (訳) 1973 教育評価法ハンドブック—教科学習の形成的評価と総括的評価— 第一法規
- R. ドラン, F. チャン, P. タミル, C. レンハート (著) 古屋光一 (監訳) 2007 理科の先生のための新しい評価方法入門—高次の学力を育てるパフォーマンス課題、その実例集 北大路書房
- 藤田育郎・池田延行・陳洋明・武田泰之 2010 走り高跳び (はさみ跳び) の目標記録への到達率からみた教科内容構成の検討: 観察的評価基準の作成と小学校高学年を対象とした縦断的実践 体育学研究 55: 539-552.
- C. ギップス (著) 鈴木秀幸 (訳) 2001 新しい評価を求めて—テスト教育の終焉— 論創社
- T. ギロビッチ (著) 守一雄・守秀子 (訳) 1993 人間この信じやすきもの 新曜社
- 池田央 1992 テストの科学—試験にかかわるすべての人に— 日本文化科学社
- 池田央 1994 現代テスト理論 朝倉書店
- 池田延行・蒲地直志 1987 体育学習における標準設定の方法に関する研究—走り高跳びについて— 体育経営学研究, 4 (1): 21-28.
- U. ナイサー (著) 富田達彦 (訳) 1989 観察された記憶 (上・下) 誠信書房
- 大村政男 1998 血液型と性格 福村出版
- Rosenthal, R., Hall, J. A., DiMatteo, M.R., Rogers, P. L. & Ar-

- cher, D. 1979 Sensitivity to nonverbal communication : The PONS test. Johns Hopkins Univ.Press.
- 佐藤浩一・入澤充・所澤潤・山口陽弘・山崎雄介・石川克博・岩澤和夫 2011 教職大学院におけるチーム・ティーチング—実践と評価、今後の課題— 群馬大学教育実践研究, 28, 241-266.
- B. D. シャクリー (著) 田中耕治 (監訳) 2001 ポートフォリオをデザインする ミネルヴァ書房
- 鈴木光太郎 2008 オオカミ少女はいなかった 新曜社
- タイラー, R. W. 1978 金子孫市監訳 現代カリキュラム研究の基礎 日本教育経営協会
- 田中耕治 (編) 2005 よくわかる教育評価 ミネルヴァ書房
- 辻井潤一・安西祐一郎 1988 機械の知 人間の知 東京大学出版界
- Wiggins, G. 1998 Educative Assessment. Jossey-Bass Pu.
- 山口陽弘 2002 試験にでる心理学〈心理測定・統計編〉 北大路書房
- 吉田正昭 1969 産業心理学 培風館
- 養老孟司・玄侑宗久 2007 脳と魂 筑摩書房

(以上 I から III は山口陽弘 担当)

IV. これからの学習評価の考え方に立った

学校現場の実践

平成22年、これからの学習評価について、目指すべき新たな方向が示された。学校現場で実践に携わる者にとっては、これまでの経緯を踏まえつつこれからの方向をしっかりとらえることが大切である。そのうえで、児童生徒の学習評価を工夫を加えながらより確かなものにするとともに、授業の改善・充実に努めなければならない。その際、I～IIIで述べられた心理学等の知見に基づく学習評価の理論を踏まえることがより効果を高めるものと考えられる。

1. 学習指導要領と学習評価

学習評価について、平成20年3月に告示された学習指導要領総則では「児童（生徒）のよい点や進捗の状況などを積極的に評価するとともに、指導の過程や成果を評価し、指導の改善を行い学習意欲の向上に生かすようにすること」としている^(注1)。

学習指導要領が改訂されるのに併行して指導要録の改訂も行われ、平成22年5月「小学校・中学校・高等学校及び特別支援学校等における児童生徒の学習評価及び指導要録の改善について（通知）」が発せられてい

る。

各学校においては、適切な教育課程を編成して目標を達成するよう教育を行うことが学習指導要領総則の冒頭に明記されている。また、児童生徒の学習状況等の評価を行うことも、学校の基本的な責務である。

したがって、教師は、児童生徒に教育活動を行うにあたっては、学習指導要領や学習評価の在り方についての趣旨や内容をしっかりと踏まえておかなければならない。特に、重要な点を下記に示す。

- ①学習指導要領では「生きる力」の育成を基本理念としているので「生きる力」がはぐくまれているかどうかをとらえることが必要である。
- ②学習指導要領に示す目標に照らしてその実現状況を見る「目標に準拠した評価(いわゆる絶対評価)」を一層重視する。
- ③P→D→C→Aのサイクルをもとに、指導と評価の一体化を図る。
- ④児童・生徒のよい点を認めて評価する。
- ⑤児童・生徒に対する評価だけでなく、指導の在り方そのものに対する評価を行う。

ところで、学習評価については、今日に至るまでにはさまざまな経緯・変遷があり、いろいろな書物にその変遷はまとめられている。その中の一冊^(注2)から上の①②に関することをかいつまんで見ておこう。

戦後間もない昭和23年度の指導要録では、評価の客観性が重視され、「相対評価」が行われることになった。

昭和46年の指導要録改訂で評定欄は「絶対評価を加味した相対評価」とされた。

昭和55年の改訂でも評定欄は「絶対評価を加味した相対評価」であったが、「観点別学習状況」欄に絶対評価が導入された。

平成3年改訂から絶対評価の「観点別学習状況」欄が基本とされたが、評定欄は相対評価であった。

「生きる力」の育成を図る評価の考え方は、平成13年度の指導要録からである。同年、「評定」欄も「相対評価」から「絶対評価」に改められた。

こうした経緯からもわかるように、「相対評価」から「目標に準拠した評価(いわゆる絶対評価)」へと移り変わり、併せて「観点別評価」が大きな位置を占めるようになったことがわかる。

2. 学力の三要素と四観点の概略

前項①で挙げたように「生きる力」の育成が、学習指導要領の基本理念となっている。特に、学力に関しては、「自分で課題を見つけ、自ら学び、自ら考え、主体的に判断し、行動し、よりよく問題を解決する資質や能力」が重要であるとしている。

こうした「生きる力」の理念は、「基礎的・基本的な知識・技能の習得を重視した上で、思考力・判断力・表現力等をはぐくむこと」を目標としている。これを、別の言い方では「我が国の子どもたちにとって課題となっている思考力・判断力・表現力等をはぐくむためには、各教科において、基礎的・基本的な知識・技能をしっかりと習得させるとともに観察・実験やレポートの作成、論述といった知識・技能を活用する学習活動を行う必要がある」とも述べている^(注3)。

さらに、学力の三要素として、以下の3点を示している^(注4)。

- ㉗基礎的・基本的な知識・技能の習得
- ㉘知識・技能を活用して課題を解決するために必要な思考力・判断力・表現力等
- ㉙学習意欲

学力をどう捉えるかについては重要な点であり、研究者によって、さまざまな学力モデルが示されている。各学校に教育行政方針を示す県や市町村の教育委員会の段階でも、国の方針に則って学力のとらえ方を示している^(注5)。

前項①②で挙げたように、学習評価は、「生きる力」がはぐくまれているかどうか、学習指導要領に示された目標・内容に照らしてその実現状況はどうか、いわゆる「目標に準拠した評価」とされている。

平成元年の学習指導要領の改訂に伴う指導要録の見直しに際して、評価の観点について、基本的には「関心・意欲・態度」「思考・判断」「技能・表現」「知識・理解」で構成し、順序もこの通りとされた。その後、平成12年12月の教育課程審議会『児童生徒の学習と教育課程の実施状況の評価の在り方について(答申)』によると、前回の方針が基本的に踏襲され、評価を行うのに当たって、「観点別学習状況の評価を各教科の基本に据え、その観点は『関心・意欲・態度』『思考・判断』『技能・表現』『知識・理解』の四観点を基本とする。」としている。これに基づいて、翌年、指導要録の改訂が行われたが、「評定」まで「目標に準拠した評価(絶

対評価)」にされただけに観点別評価のもつ意味は従来以上に重みを増すことになった^(注6)。

さて、現在は、IVの1で述べたように、改訂された学習指導要領と指導要録にもとづいて指導と評価の一体化を図っていくことになる。

評価の観点は引き続き四観点ではあるが、「基礎的・基本的な知識・技能の習得とこれらを活用する思考力・判断力・表現力等をいわば車の両輪として相互に関連させながら伸ばしていくとともに、学習意欲の向上を図る」という改訂の趣旨を生かすとして以下の四観点到に整理がなされた。

- ①「関心・意欲・態度」
- ②「思考・判断・表現」
- ③「技能」
- ④「知識・理解」

したがって、評価の四観点と上述した学力の三要素との関係を、教科によって違いはあるもののおおむねつぎのように整理できるとしている。

◆観点④「知識・理解」と③「技能」が、学力㉗基礎的・基本的な知識・技能の習得、

◆観点②「思考・判断・表現」が、学力㉘知識・技能を活用して課題を解決するために必要な思考力・判断力・表現力等、

◆観点①「関心・意欲・態度」が、学力㉙学習意欲、学習指導要領の改訂に伴って、冒頭に掲げた「小学校・中学校・高等学校及び特別支援学校等における児童生徒の学習評価及び指導要録の改善について(通知)」に、学習評価における整理された四観点が示されている。学習指導を行うのも、学習評価を行うのも、学校現場の教師一人一人であり、それは日々の教育活動の中にあることを再確認したい。

3. 学校現場では

現場の教師は学習指導要領やその解説等を参照し、趣旨・内容を受け止めた授業をしようとするがそれを評価にまで結びつけていくことやその結果を授業改善に生かそうというところまでなかなか行き着かない傾向が見られる^(注7)。評価に関する諸資料が提示されているにもかかわらず意外と参照されていない様子も見られる。日常の多忙感がそうさせているとも推察される。

たとえば、既述の「小学校・中学校・高等学校及び

特別支援学校等における児童生徒の学習評価及び指導要録の改善について（通知）」、これに先立つ「児童生徒の学習評価の在り方（報告）」、さらにこの報告の前提となった「学習指導と学習評価に対する意識調査報告書」など、学習評価の改善につながる一連の資料がある。また、国立教育政策研究所が発行している「評価規準作成のための参考資料」（平成22年11月）、「評価方法等の工夫改善のための参考資料」（平成23年3月）も評価を具体化するうえで必携すべきものである。

加えて、群馬県教育委員会でも「これからの学習評価について」（平成22年8月）を作成している。

しかし、最終的には各学校が評価規準を設定しなければならないのであり、各学校、教師一人一人がこれらの諸資料を十分に参照した上で、主体的な取り組みをすることが重要である。

V. 真正の評価による授業の改善・充実をめざして

Iで提起された真正の評価をめざした授業を考え実践することは、指導と評価の一体化を図るうえで大切なことと思う。しかし、そう簡単なことではない。日頃行っている授業を真正の評価という視点から見直し、改善・充実に努めるという方が、取り組みやすく現実的かも知れない。

ところで、現場の教師は、学習指導要領の目標・内容を踏まえ、教科書を使用して授業を進めている。上述したように、評価についても学習指導要領と表裏一体となった形でその在り方が示されている。教師が指導し評価する際に、学習指導要領に示された目標や学習内容が児童生徒の身に付いているかどうか、より妥当性・信頼性のある適切な評価に心がけなければならない。

日頃取り組んでいる授業を真正の評価という視点をもって改善・充実させる一つの手立てとして、テスト問題・結果の活用があげられる。

1. テスト問題・結果の活用

テストは、児童生徒の学習状況を知るとともに、教師の授業を評価するものの一つでもある。平成19年度以来、毎年、実施されてきた全国学力・学習状況調査は真正の評価を考える際のヒントを投げかけている。テスト結果の数値や成績の順位に関心が向きがちであ

るが、結果を踏まえて授業の改善・充実にあたることこそが重要である。このテスト問題では、学習指導要領で求める学力に対しての実状を問うている。周知のようにA問題・B問題の大きく2つの区分で実施されているが、A問題が「主として『知識』に関する問題」であるのに対して、B問題は「主として『活用』に関する問題」としている^(注8)。B問題において、正答率が低いことが話題になっている。

①国語 学級会における話し合い活動

平成21年度の小学校国語B問題²に「川口さんの学級では、家族の一員としてできることを考えるために、家族での過ごし方について調べました。次は、川口さんのグループが集めた資料をもとにした話し合いの様子の一部です。よく読んであとの問いに答えましょう。」という問題がある^(注9)。

この「話し合いの一部」には司会・川口・松山・村田の発言が記され、あなたの発表として空欄の□があり、後の設問で解答者もこの話し合いに加わって自分の考えを述べることになる。司会は資料をもとにして、「家の中のそうじや整とん」について話し合うことを述べ、資料を見て分かったことや考えたことの発表を求めている。資料とは「家の中のそうじや整とんをする小学校六年生の割合」という統計資料である。

設問1は、話し合いの中で出された川口・松山・村田の3名の意見を二つの立場、すなわち、「そうじや整とんによく取り組んでいる」とするAと「そうじや整とんにあまり取り組んでいない」とするBとに分ける。

設問2は、「あなたは、……」と問いかけ解答者に二つの意見のうちAの立場からの発表を求めるとともに、2つの条件を付している。条件とは、①資料中の統計は平成16年度・平成17年度の2年分が掲げられているが平成17年分を取り上げること、②六十字以上八十字以内にまとめて書くことである。

この問題の趣旨は学習指導要領第五学年及び第六学年のA「話すこと・聞くこと」で示された「互いの立場や意図をはっきりさせながら、計画的に話し合うこと。」ができるかどうかを見るのがねらいという。

設問1では「話し手の立場や意図をとらえて聞くこと」設問2では「自分の立場や意図をはっきりさせながら話し合う」ことができるかどうか問われている。

これは、五年及び六年の目標として掲げられている「目的や意図に応じ、考えたことや伝えたいことなど

について、的確に話す能力、相手の意図をつかみながら聞く能力、計画的に話し合う能力を身に付けさせるとともに、「……」に照らした達成状況を見ようとするものといえよう。

この問題の正答率は設問1が75.6%、設問2が25.9%であったとされている。このような問題をヒントに授業の在り方を改善していこうとすることが大切と考える。「話す」「聞く」「話し合う」という活動を低学年の段階から組織的・計画的に進めていく必要があるだろう。

上で見てきたように、この問題の場面設定でも二つの意見があることを明確にして話し合いを進めている。日常の授業の中で、さまざまな立場の意見があることをとらえる力を培う必要があるだろう。ある議題に賛成と反対の立場がある場合に、自分がいずれかの立場に立った時、賛成の立場から考えたり、反対の立場から考えたりするなど両者の立場についての理解・認識を深める段階が大切である。そのうえで最終的に自分の立場や意見をまとめていく。そうした実際的な経験を積む場を授業の中で意図的に設定することが必要である。もう一つ、この問題では資料で示された数値をもとにした話し合いをしている。単なる思いつきや感覚で意見を述べているのではなく、統計表を根拠にして自分の意見を述べている。話し合う議題に応じて必要な図表やグラフなどの資料を集め用意して根拠にもとづいた話し合いが求められている。これも日頃の授業の中でそのような実体験がなければ容易には身に付かないであろう(注10)。

この問題でもう一つ着目したいのは「自分の考えの形成」という点である。B問題では「自分の考え」を所定の字数で記述することが求められるケースが多い。自分の意見や考えがもてるということはたいへん重要なことであり、これも段階を追ってそうした力をはぐくむように授業の中で計画的に取り組んでいくことが必要である。

②算数 平行四辺形の面積

平成19年度の小学校算数B問題[5]の平行四辺形の面積の問題は、その後さまざまな書物等でも取り上げられて広く知られるところとなった。

とくに、(3)「ひろしさんの家の近くに東公園があります。東公園と中央公園の面積では、どちらの方が広いですか。答えを書きましょう。また、そのわけを

言葉や式を使って書きましょう。」は正答率が18.2%と低いものであったという(注11)。

この問題では、ひろしさんの家の近くの様子が地図で示され、その中に二つの公園、すなわち、東公園と中央公園が図示されている。道路と道路が「平行」や「垂直」になっている等の説明がなされているとともに、道路と道路の間の距離が図中に数値として6箇所記入されている。

まず、このような図や条件文・数値等から2つの公園がそれぞれ長方形と平行四辺形であることを読み取らなければならない。そのうえで、両者の面積を出すのに必要な数値を見つけ出して計算し、その結果を比べて広い方を答える。問題を解くのに必要のない情報も入っており、しかも、平行四辺形の中央公園は斜辺にあたる数値が公園に近い場所に表示されているので、これに惑わされて誤答してしまったケースが多い。図中の「高さ」にあたる数値を見つけて正しく計算しなければならない。ここがひとつのポイントで、公式を機械的に暗記しているだけの浅い理解ではなく、深い意味理解がなされているかが問われている(注12)。

もう一方の長方形の東公園についても必要な数値を見つけて面積を求め、どちらが広いかを判断する。問いでは判断した「わけを言葉や式で説明すること」も求めている。これは、「思考・判断・表現」という新しい観点の考え方が明確にわかる問いかけでもある。

したがって、平行四辺形についての授業の段階において、意味理解を深める課題設定や問いかけをして、「思考・判断・表現」する力をはたらかせる場をつくっていくことが大切であると思う。まさに、こうしたテスト問題から授業の中でおさえなければならないポイントが浮かびあがってくる。その一つが平行四辺形の「高さ」のとらえ方であり、ここに指導上の弱さがあると見た場合には強化策を打ち出しているケースもある。調査結果が良好とされた秋田県においても「高さ」の指導に力を注ぐ必要があるとして指導資料を作成して周知している(注13)。

また、学習指導要領では、第五学年で平行四辺形の面積の求め方を学習することになっているが、その際、算数的活動として「具体物を用いたり、言葉、数、式、図を用いたりして考え、説明する活動」を求めている(注14)。

この問題は、まさに学習指導要領で求めている学習

内容が身に付いているかどうかを評価している。こうした「B問題こそ、『真正の評価』論に基づく『パフォーマンス評価』の一例といえます。」^(注15)という考え方も傾聴に値する。

全国学力学習状況調査のように、学習指導要領の目標や内容等との関係を精査して作成された評価問題は児童生徒の解答をしっかりと分析してみることが大切と思う。そして、授業そのものが学習指導要領の目標や内容を十分踏まえた適切なものになっているかどうかを教師自身が振り返ることで指導に生かすべきものと考ええる。

そうした点で、国の機関の発行物ばかりでなく県や市町村教育委員会で作成している授業改善の手引きなども関連させてみると有機的なつながりをもってとらえることができる。ちなみに、前橋市が作成した資料の一例をあげると、「4年間の全国学力・学習状況調査の問題を改めて見直すと、児童生徒に身に付けさせたい力が見えてきます。」^(注16)とある。こうした冊子を参照しながら学習指導上のポイントを改めて確認するとともに単元計画や年間指導計画等にも反映させる必要があるだろう。

実践の場において、全国学力・学習状況調査というひとつの学習評価を授業改善に生かすことは、IのIで論じられているエバリュエーションとアセスメントの関係を具体的にしたものともいえよう。

2. パフォーマンス評価

上述の算数や国語のB問題のように「思考力・判断力・表現力」を評価しようとするパフォーマンス評価とされる方法を考える傾向が強まっている^(注17)。パフォーマンス評価というと完成作品や口頭発表・実技等を思い浮かべるが、先ほどの学級会の場面で自分の意見を一定の条件のもと、所定の字数で述べることや四角形の面積の比較で広いと判断したわけを言葉や式で述べることも完成作品の一つともいえよう。その際、児童の解答を予め予測し評価規準を設定しておかなければならない。3段階なら「おおむね満足」「十分満足」「不十分」ということになるであろう。点数化する場合もあるが、いわば、ルーブリックの作成である。

B問題のような真正の評価を意識しながら授業改善を進めていくとパフォーマンス評価を取り入れる場面が増えていくと思われるが併せてルーブリックづくり

も行わざるを得なくなる。

一方で学校現場の毎日は授業の充実に力を注いでいるのはもちろんであるがその他の多様な業務もあり忙殺されているのが現実である。限られた時間の中で必要なことに時間をかけながらも効率よく進めて行かなければならない。授業改善が効果をあげるためにもパフォーマンス評価やルーブリックも過度な負担がないようにしていく工夫が必要と考える。作品や口頭・実演という方法もあるが、全国学力調査のようにテスト問題で評価することも一方法である。

3. 妥当性・信頼性の向上に向けて

学校や教師は児童生徒の学習評価を実施しているがその妥当性・信頼性を高めることが求められている。このことは、実際に行われている学習評価の妥当性・信頼性が十分でない面があるのではないかということの裏返しでもある。まずは、妥当性を確保することが大切であろう。そのためには、①「評価結果と評価しようとした目標の間に関連性があること」②「評価方法が評価の対象である資質や能力を適切に把握するものとしてふさわしいものであること」が必要とされている^(注18)。学習の目標・内容と評価の対象を明確にするとともに、評価規準を設定し、適切な評価方法を選択する。さらに、学習評価をもとに授業の改善・充実に努めることが求められている。授業研究・校内研修等を通して教師の授業力を高めることこそが重要であり、妥当性・信頼性の向上につながるであろう。

上述IIの2で「クレバーハンスの馬」、IIの3で「ピグマリオン効果、ハロー効果」について述べられている。

学校現場ではこれら指摘される状況が起きている場合があるかもしれない。たとえば、授業中での発問と応答の関係が知らずのうちに誘導的になりながら進行している場合などである。教師自身がそのような事例があるということ踏まえて指導に当たることが必要だろう。授業で発問を考える際に、児童生徒の反応・解答を予測し、クレバーハンスの馬のように教師の顔色を見て反応したものにならないようにすることも考えなければならない。また、実際の児童生徒の反応にもそうした要素が加味されることもあるかもしれないので、その場合には適切な問い返しができる力量を培いたいものである。

ところで、教師からの言葉かけがピグマリオン効果をおこし、児童生徒の力が結果的に大きく伸びたということであればプラスの効果といえよう。

一方で、児童生徒の一方的な評価を他の側面に及ぼして適用して評価するなどマイナスのハロー効果であり、妥当性を欠いた評価ということになる。

学校現場では、日々いろいろな現象が起きている。自分たちの教育活動の背景にある教育理論や、国・県・市町村などの施策も承知したうえで、冷静な目で振り返り、客観的に判断・対処できることが学習評価の妥当性・信頼性の向上にもつながると考える。

【注】

- (注1) 文部科学省『学習指導要領 総則』平成20年3月「第4指導計画の作成等に当たって配慮すべき事項」
- (注2) 田中耕治編『よくわかる教育評価』2005年ミネルヴァ書房 IX指導要録138～149頁(樋口太郎)特に、表30「戦後児童指導要録の特徴」において、戦後の学習評価の変遷がコンパクトにまとめられている。
- (注3) 中央教育審議会『幼稚園、小学校、中学校、高等学校及び特別支援学校の学習指導要領の改善について(答申)』平成20年1月22～23頁
- (注4) 学校教育法第30条第2項 及び 文部科学省『学習指導要領 総則』平成20年3月
- (注5) ちなみに、群馬県教育委員会では、平成22年度教育行政方針の中で「基礎・基本の確実な習得のために」を標題に掲げるとともに、学力の諸要素の関係を図示している。
- (注6) 観点別評価の重要性について北尾倫彦・金子守編『観点別評価ハンドブック 中学校編』2003年6月 図書文化社 「第一に、学習の成果を全体的にとらえる必要があるといえる。(中略) 第二に、自ら学び自ら考える力を育成しようとするという学校教育を実現するためには、意欲や思考力などを重視した評価を行う必要がある。(中略) 第三には、学習のプロセスを重視し、指導と評価の一体化を図るためには観点別評価が重視される必要がある。」8～9頁
- (注7) 中央教育審議会初等中等教育分科会教育課程部会『児童生徒の学習評価の在り方(報告)』平成22年3月の中においても「現在の学習評価については、負担感や授業改善に関して課題があると考えられる。」と指摘されている。
- (注8) 『全国学力学力学習状況調査 解説資料』(国立教育政策研究所)による。A問題は「①身につけておかなければ後の学年等の学習内容に影響を及ぼす内容②実生活において不可欠であり、常に活用できることになっていることが望ましい知識・技能」としている。B問題は①知

識・技能等を実生活の様々な場面に活用する力②様々な課題解決のための構想を立て、実践し評価・改善する力にかかわる内容」としている。

- (注9) 国立教育政策研究所教育課程研究センター『平成21年度全国学力・学習状況調査 解説資料 小学校 国語』平成19年5月
- (注10) この問題をもとにした授業改善の提案として国立教育政策研究所『全国学力学習状況調査小学校の結果を踏まえた授業アイデア例』平成21年8月がある。3頁 http://www.nier.go.jp/09jugyourei/21_shou_jugyou_idea_houkoku.pdf
- (注11) 国立教育政策研究所教育課程研究センター『平成19年度全国学力・学習状況調査 解説資料 小学校 算数』平成19年5月
- (注12) 田中耕治「学力調査に見る日本の子どもの特徴と弱点—算数学力に焦点を当てて」『児童心理』2009年2月NO.890 金子書房 20～21頁
- (注13) 秋田県教育委員会『全国学力・学習状況調査算数・数学(小6・中3) 授業改善に向けて』による。全国の正答率を上回っていたものの課題が見られるとして、各学校に対し「評価問題として、情報過多の問題や高さを確認する問題を出題し、定着を図るなど」の提言を行い、「即、その後の授業から実践」と述べている。
- (注14) 文部科学省『小学校学習指導要領解説 算数編』平成20年8月 東洋館出版社 149～151頁において指導のポイントを詳述している。
- (注15) 田中耕治『新しい「評価のあり方」を拓く—「目標に準拠した評価」のこれまでとこれから—』2010年11月日本標準ブックレットNO.12 40頁
- (注16) 前橋市教育委員会『全国学力・学習状況調査に算数・数学科で身に付けさせたい力』平成22年作成。4年間の全国調査問題を児童生徒に身に付けさせたい視点で整理して学習指導のポイントを示している。
- (注17) 中央教育審議会初等中等教育分科会教育課程部会『児童生徒の学習評価の在り方について(報告)』平成22年3月24日の中で「評価規準や評価方法については、近年諸外国においても様々な研究や取組が行われており、例えば、『思考・判断・表現』に関する評価規準としては、学年等ごとに細分化したものを定めるではなく……』という説明のうち、研究や取組の例として注釈で、パフォーマンス評価について、次のように述べている点が注目される。「思考力・判断力・表現力等を評価するに当たって、『パフォーマンス評価』に取り組んでいる例も見られる。パフォーマンス評価とは……」
- (注18) 中央教育審議会初等中等教育分科会教育課程部会『児童生徒の学習評価の在り方について(報告)』平成22年3月24日の中で、「学習評価の『妥当性』は評価結果が評価の対象である資質や能力を適切に反映しているものであることを示す概念として用いている。」としている。「学

校や教師は、評価の実施者として、個々の児童生徒の学習評価に関する妥当性・信頼性を高め説明責任を果たすとともに……」とか、「目標に準拠した評価の妥当性、信頼性等を確保していくためには、学校おける組織的な取組の充実や保護者の理解の促進を更に図っていくことが重要である。」など、妥当性・信頼性に関する記載が見られる。

工藤文三「これからの学習評価の考え方と評価方法の工夫改善」『初等教育資料』平成23年1月号 NO.869 東洋館出版社

(以上、IVからV 石川克博 担当)

(やまぐち あきひろ・いしかわ かつひろ)